A novel data management platform to improve image-guided precision preclinical biological research

Citation for published version (APA):

Persoon, L., van Hoof, S., van der Kruijssen, F., Granton, P., Rivero, A. S., Beunk, H., Dubois, L., Doosje, J-W., & Verhaegen, F. (2019). A novel data management platform to improve image-guided precision preclinical biological research. *British Journal of Radiology*, *92*(1095), [20180455]. https://doi.org/10.1259/bjr.20180455

Document status and date:
Published: 01/01/2019

DOI:
10.1259/bjr.20180455

Document Version:
Publisher's PDF, also known as Version of record

Document license:
Taverne

BJR

https://doi.org/10.1259/bjr.20180455

Received: 18 May 2018 | Revised: 16 September 2018 | Accepted: 19 September 2018

Cite this article as:
Persoon L, Hoof S, van der Kruijssen F, Granton P, Sanchez Rivero A, Beunk H, et al. A novel data management platform to improve image-guided precision preclinical biological research. *Br J Radiol* 2019; **92**: 20180455.

SMALL ANIMAL IGRT SPECIAL FEATURE: FULL PAPER

A novel data management platform to improve image-guided precision preclinical biological research

[1]**LUCAS PERSOON**, PhD, [2]**STEFAN VAN HOOF**, MSc, [1]**FRANK VAN DER KRUIJSSEN**, MSc, [2]**PATRICK GRANTON**, PhD, [1]**ANDREA SANCHEZ RIVERO**, BSc, [1]**HAROLD BEUNK**, BSc, [3]**LUDWIG DUBOIS**, PhD, [1]**JAN-WILLEM DOOSJE**, BSc and [2,3]**FRANK VERHAEGEN**, PhD

[1]Healthcare department, ICT Group N.V, Eindhoven, Netherlands
[2]Smart Scientific Solutions B.V, Maastricht, Netherlands
[3]Department of Radiation Oncology (MAASTRO), GROW–School for Oncology and Developmental Biology, Maastricht University Medical Centre, Maastricht, Netherlands

Address correspondence to: Dr Lucas Persoon
E-mail: *lucas.persoon@ict.nl*

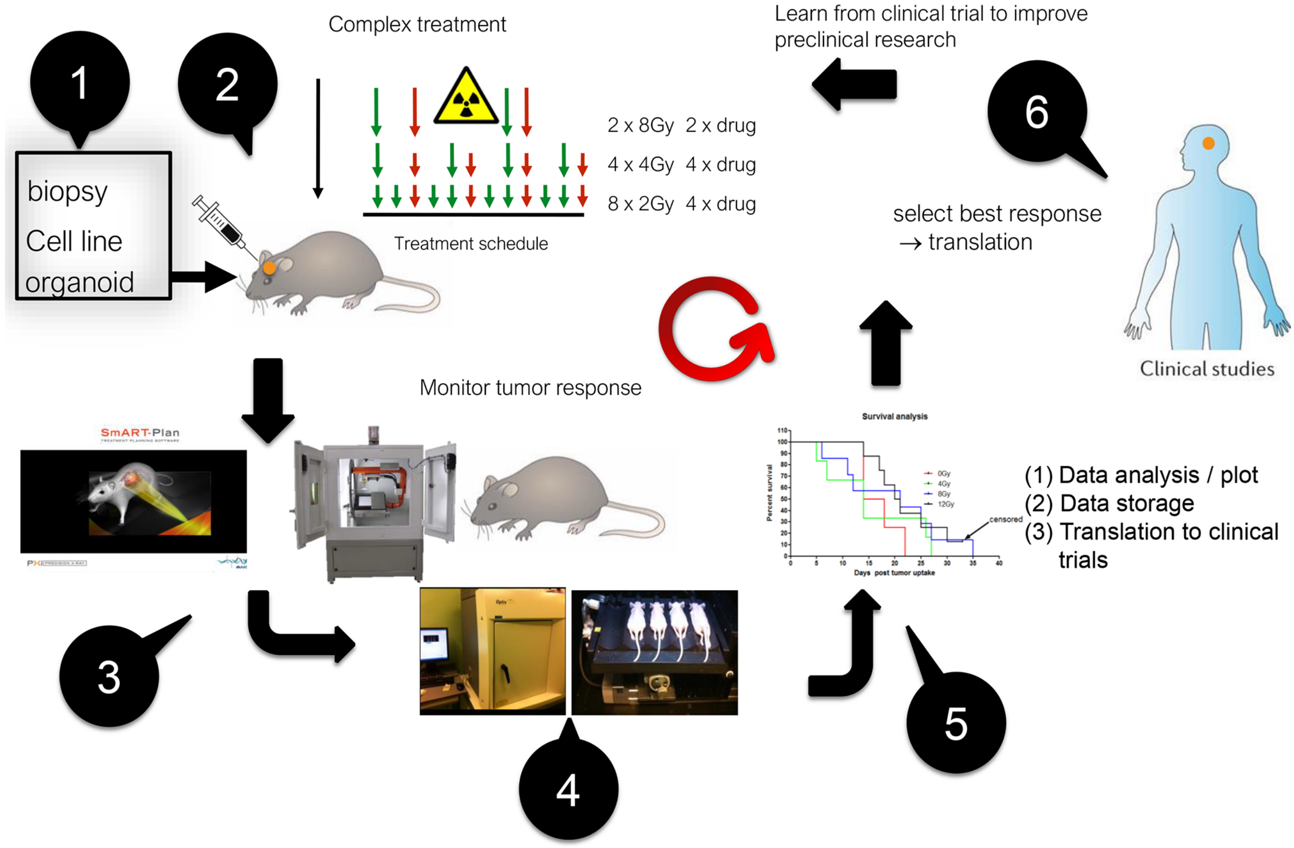
Objective: Preclinical biological research is mandatory for developing new drugs to investigate the toxicity and efficacy of the drug. In this paper, the focus is on radiobiological research as an example of advanced preclinical biological research. In radiobiology, recent technological advances have produced novel research platforms which can precisely irradiate targets in animals and use advanced onboard image-guidance, mimicking the clinical radiotherapy environment. These platforms greatly facilitate complex research combining several agents simultaneously (in our example, radiation and non-radiation agents). Since these modern platform can produce a large amount of wide-ranging data, one of the main impediments in preclinical research platforms is a proper data management system for preclinical studies.
Methods: A preclinical data management system, inspired by current radiotherapy clinical data management systems was designed. The system was designed with InterSystems technology, *i.e.* a programmable Enterprise Service Bus solution. New DICOM animal imaging standards are used such as DICOM suppl. 187 for storing small animal acquisition context and the DICOM second generation course model.
Results: A small animal big data warehouse environment for research is designed to work with modern image-guided precision research platforms. Its modular design includes (1) a study workflow manager, (2) a data manager, and (3) a storage manager. The system provides interfaces to, *e.g.* preclinical treatment planning systems and data analysis plug-ins, and guides the user efficiently through the many steps involved in preclinical research. The system manages various data source locations, and arranges access to the data centrally.
Conclusion: A novel preclinical data management system can be designed to improve preclinical workflow, facilitate data exchange between researchers, and support translation to clinical trials.
Advances in knowledge: A preclinical data management system such as the one proposed here would greatly benefit preparation, execution and analysis of biological experiments, and will eventually facilitate translation to clinical trials.

INTRODUCTION

In radiotherapy clinical practice it is common to handle large amounts of patient data with efficient data management systems. Examples hereof are PACS (Picture Archiving and Communication System)[1] treatment planning systems that include radiobiological modelling and can transfer treatment instructions to irradiation equipment (*e.g.* linear accelerators), DICOM (Digital Imaging and Communications in Medicine)[2] standards for images, treatment plans, dose distributions and structure delineations, which facilitate easy data processing and data transfer in the clinic and between different clinics. In particular, the DICOM standard is well supported by a wide range of imaging and treatment device manufacturers and healthcare IT support services.

In the upcoming field of image-guided precision radiation research, the aim is to empower preclinical animal trials of combinations of radiation with other agents such as drugs, chemotherapy, immunotherapy etc. The novel radiation research platforms[3,4] that became available recently are heavily used for cancer research, but could also be used for other diseases in, *e.g.* neurology, cardiology etc. Besides using radiation beams to destroy cancer cells, these platforms can be used to introduce selective damage in various organs.[5] Figure 1 shows a typical workflow of modern

Figure 1. Example of a typical workflow in modern preclinical image-guided precision radiotherapy experiments. Starting at the top left and then counter-clockwise: (1) creation of a tumor model in a small animal, (2) Experimental schedule with several arms, (3, 4) planning and execution of image-guided precision irradiation, (5) data collection and analysis, and (6) translation into the clinic.



preclinical cancer research: starting from a biological model (cell culture, organoid, animal), response of normal or cancerous tissue to radiation combined with other agents is studied.[6–8] While there has been much recent progress in developing tumor models (*e.g.* orthotopic) and precision irradiation combined with imaging,[9,10] the current bottleneck is data storage, retrieval, communication and analysis for these experiments that may generate large amounts of data, *e.g.* in multimodality imaging longitudinal studies. Other aspects such as defining research protocols, and efficient time-management of the workflow of experiments, could also be substantially improved by borrowing and adapting methods from clinical practice. The latter is often far ahead of data handling methods currently employed in preclinical experiments.

In this paper we present a proposed design of a data management platform to improve image-guided precision preclinical research, and to empower big data analysis. This platform is inspired by current and near-future clinical methods, but also contains some aspects specific to animal studies.

DESIGN OF THE DATA MANAGEMENT SYSTEM

Requirements for the data management platform

Among the sources of data that may need to be handled in modern preclinical experiments are (non-exhaustive list): (1) study set-up, (2) treatment schedules, (3) radiation treatment plans, (4) various sources of imaging data (MRI, PET, SPECT, CT, dual-energy CT, multi energy CT, cone beam CT, digital pathology, histology and other microscopy images, (5) radiomics data, (6) genomics data, (7) data analysis files and reports (text and graphics) etc. Imaging-based cancer research is in the early stages of a revolution towards digitization, which also requires more advanced tools for data standardization and management. In our example, several commercially available radiation research platforms [notably XRAD-SmART from PXi (North Branford, CT), and SARRP from XStrahl Ltd (Camberley, UK)] currently support the DICOM standard for imaging and radiotherapy DICOM-RT objects (RT Image, RT Structure Set, RT Plan, RT Dose). Recently DICOM working Group 30 issued DICOM standard supplement 187[11,12] for the acquisition of preclinical small animal imaging data. This is about 25 years after the introduction of the clinical DICOM imaging standard and about 20 years after adding the DICOM-RT objects.[2] It is expected that this will greatly facilitate preclinical data management.

However, to ensure efficient handling of the vast amount of data that modern preclinical trials can generate, a data management platform should ideally be capable of:

(1) managing the workflow of entire animal experiments (preparation, execution, analysis)

(2) handling large volumes of data from many different sources stored at different locations
(3) integrating with preclinical systems (*e.g.* treatment planning systems)
(4) reporting of the study results
(5) handling complex queries and allow/create different views or representations of the (experimental) data
(6) supporting plug-ins for outcome prediction models
(7) supporting inter- and intra institutional communication between data management platforms
(8) translating preclinical findings into human clinical trials.

While many of these requirements are similar to the requirements for clinical operation, adaptive radiotherapy and clinical trials, a number of elements are specific to animal studies. The most important differences in requirements are related to the underlying data model, which needs to be extended. For clinical use a patient oriented DICOM second generation RT-Course record model is desirable to record/orchestrate all data acquired, while for small animal research an animal cohort (collection of study subjects) oriented data model is mandatory. Besides the registration of acquired data in its context (workflow), information and context about the different study arms (a subset of experiments) and the cohort needs to be captured. Therefore, the data model has to be extended with study arm and cohort specific data and information. This extension of the data model is critical for research purposes while in a clinical setting data captured in its context is the most important aspect. Other differences are at the level of the interactions of the platform with the many different imaging and other devices encountered in animal research, and for the data generated from *ex-vivo* tissue analysis.

General description of the data management platform

Figure 2 shows two representations of an animal data management platform. The left panel shows the different levels where the system can support research. At the lowest level the system manages the pre-clinical operation, it supports the research team in their daily activities: it helps to execute experiments and trials efficiently, it facilitates daily data handling, and it provides Application Programming Interfaces to other systems (*e.g.* radiation treatment planning system) and research software (*e.g.* MevisLab image analysis modules).

At a higher level the management of the preclinical workflow and data takes place. Here the research cohorts and studies are managed, *e.g.* planning and managing experiments including, *e.g.* a link to an electronic calendar to alert researchers to their tasks at hand. Templates for workflow description are provided. Consistency of data reporting is also assured at this level.

At the top level tools for big-data analysis, mining and trend analysis are provided. This is also the level where the translation of studies into clinical trials is facilitated. This is arguably the most challenging part and where most of the knowledge is currently lacking. Rules to translate preclinical findings into human trial protocols can be provided at this level, taking into account that these rules may still need to be discovered.

The right panel of Figure 2 provides a general view of the management stages (workflow, data and storage) and the interactions necessary to support preclinical research. Workflow management is necessary to be able to perform consistently the research and provide support on which activity to execute and how to do it. Workflow management has a close interaction with data management as results/data (*e.g.* document, DICOM files, reports, images) produced throughout the workflow have to be managed within the context of the creation and life-cycle. Workflow manages the activities which need to be performed and data management controls the conditions of how data was produced within the workflow context. Storage management is concerned with where the produced data actually resides (*e.g.* PACS, file-share) within the research department. Storage management essentially keeps data at the source and prevents undesirable data duplication with the risk of inconsistent data. The workflow management guides the process of producing data and keeps the data record consistent. Data also produces workflow: data triggers research and other activities such as storage and analysis, and the workflow guides the user through the next steps in the process in the correct sequence.

Figure 2. Left panel: This shows the various levels where a data management system can support preclinical research. Right panel: Schematic overview of how all stages of data production, storage, management and workflow interact.

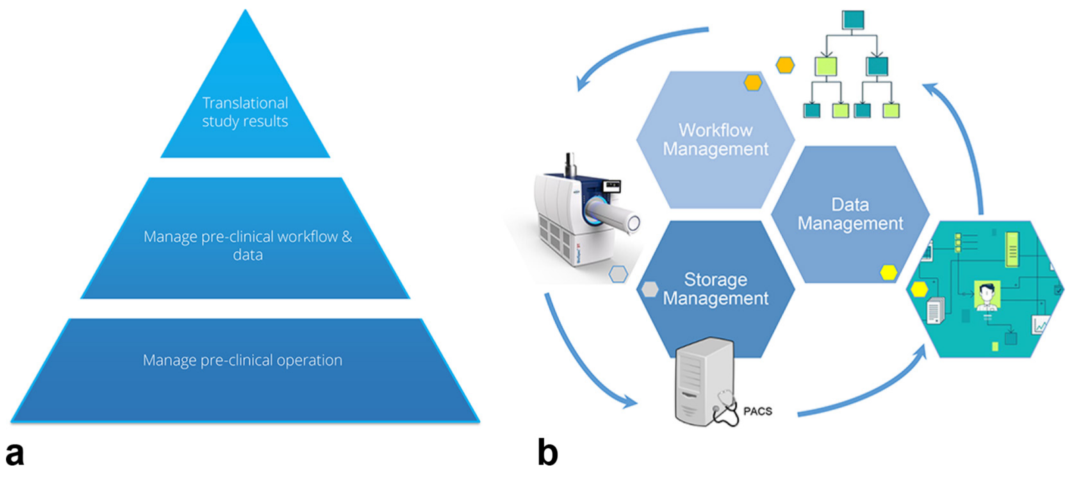
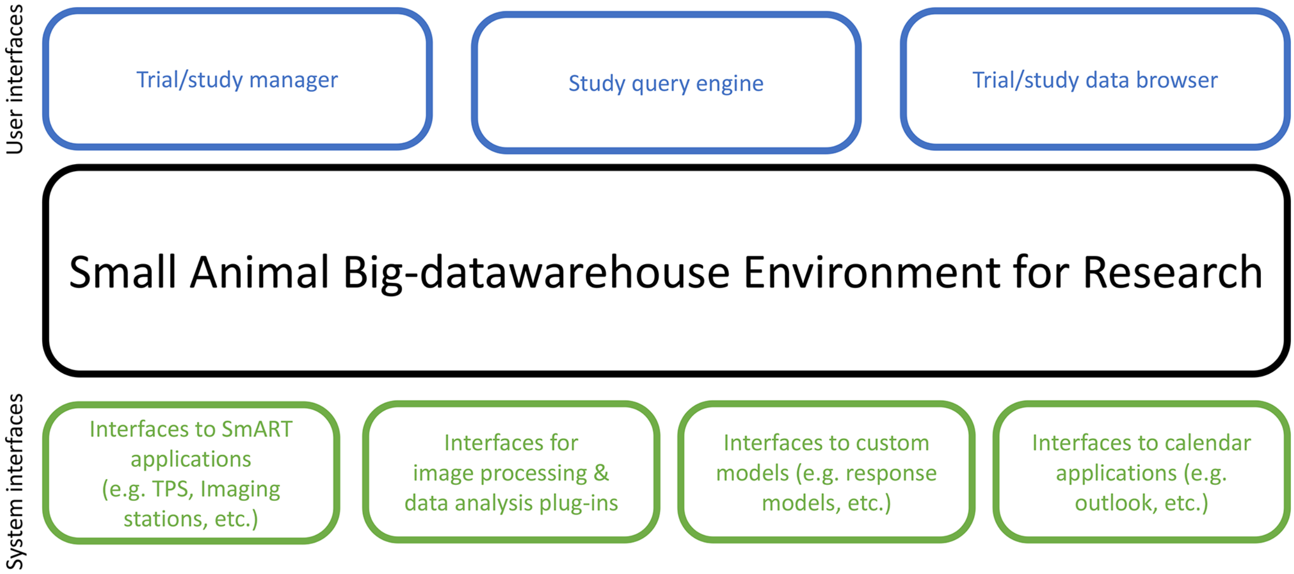


Figure 3. Data management system design.



RESULTS

Data management system

A design of a data management system for preclinical research is shown in Figure 3. Central to the system is an intelligent Small Animal Big-data warehouse Environment for Research (SABER). A commercial enterprise service bus from InterSystems (Cambridge, MA) was used. This system contains an ultra-high performance database (InterSystems Caché) and interconnectivity for most medical protocols[13] (*e.g.* DICOM, HL7, FHIR, etc.) and non-medical interface protocols (*e.g.* SOAP, REST). The InterSystems Caché is by nature an unstructured database but capable of supplying structured views at, *e.g.* the study execution side from a cohort/study/animal perspective while simultaneously being capable of near real-time data views and analysis by providing connectors to big data analysis tools, *e.g.* Apache-SPARK. This can be utilized for big-data purposes. SABER provides various interfaces to, *e.g.* animal irradiators, radiation treatment planning systems (*e.g.* SmART-ATP, SmART Scientific Solutions BV, Maastricht, Netherlands)[14] and communication to image acquisition stations. It also provides standardized or proprietary interfaces for image processing and third-party image analysis software (plug-ins). There are many different options available in preclinical research such as P-MOD, MevisLab, MiM, CERR, Imalytics, VivoQuant, 3DSlicer, Osirix, Amide, AsiPro, to name just a few. These packages allow a range of data and graphical analysis from very simple to highly complex such as data mining and radiomics image feature analysis.

SABER also provides interfaces to various custom models (*e.g.* dose response models, prediction models) and to an application programming interface (API) that allows developing one's own models (*e.g.* in MATLAB, Mathworks, Natick, MA or Python, Python Software Foundation, …).

In the top layer of Figure 3 various user interfaces to handle the study which require user interaction are depicted. The Trial/study manager serves to formulate the research questions, set up the study parameters, the cohort(s), define all the parts of the study (imaging, irradiation, data analysis), ensures that no study arms are forgotten; it defines the whole workflow (see also the example in Section 3.3). This is one of the most powerful elements in the data management system. A Trial/study data browser allows quick overview and navigation of the acquired information and the details of a single study. Also very important is the Study query engine which supplies a view across multiple studies. This uses intelligent and efficient software to extract information about the cohorts using a visual query builder, based on the InterSystems Query language. A large advantage is that structured data can be retrieved much faster.

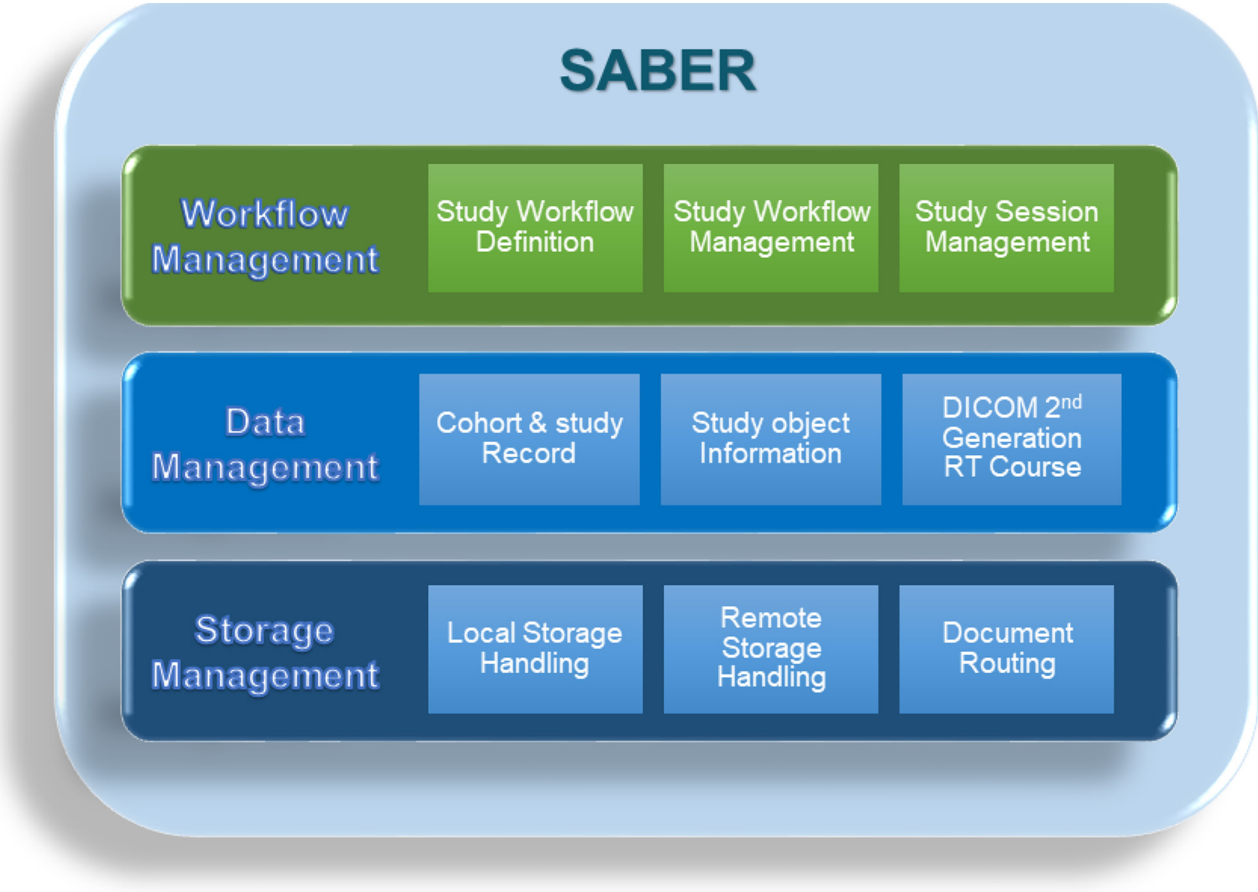
Small animal Big-Data warehouse

The core of the data management system is the SABER, a "black box" that the user does not need to know the details about, consisting of three layers: a workflow management layer, data management layer and storage layer (Figure 4).

The workflow management layer records the study workflow definitions, the model/templates of the workflow to be followed for a research project and it records the state of the active workflows, determining whether activities in the workflow are due or late depending on whether pre- and post-conditions are met etc. to enable/disable next steps and inform the user about workflow status. Workflow flexibility is key, and changes can be made at all times during all phases of the experiments.

The data management layer keeps track of cohort and study records, manages the study object information and employs the DICOM second generation radiotherapy course record (the course record is the basic structured data record of the acquired information. It is the data-model used to capture the information/data for each executed workflow). The latter is the orchestrating data management framework for each individual workflow. The course record provides a means to put the produced study data in its context, in our example here a radiotherapy small animal study/treatment. The data management layer (Figure 4) also

Figure 4. SABER design. SABER, Small Animal Big-data warehouse Environment for Research.



allows data files in many different formats to be handled (*e.g.* csv, pdf, doc, xls, …).

The storage management layer takes care of the actual storage and handling of the data. It manages by default that all information local and external about the studies is stored securely (including a backup system). SABER relies on data provided by different sources, therefore to prevent data being duplicated the storage philosophy of SABER is to let data remain at the source where it was generated. SABER keeps only a reference/record of where the data resides. For example images stored in a PACS are not copied but what is stored is a record of the image and its location, in this case the PACS storage device, reducing the amount of storage needed but keeping a consistent record. Furthermore, the storage layer is responsible for routing documents from one source to another. Documents are defined as data produced during execution of a task in the workflow.

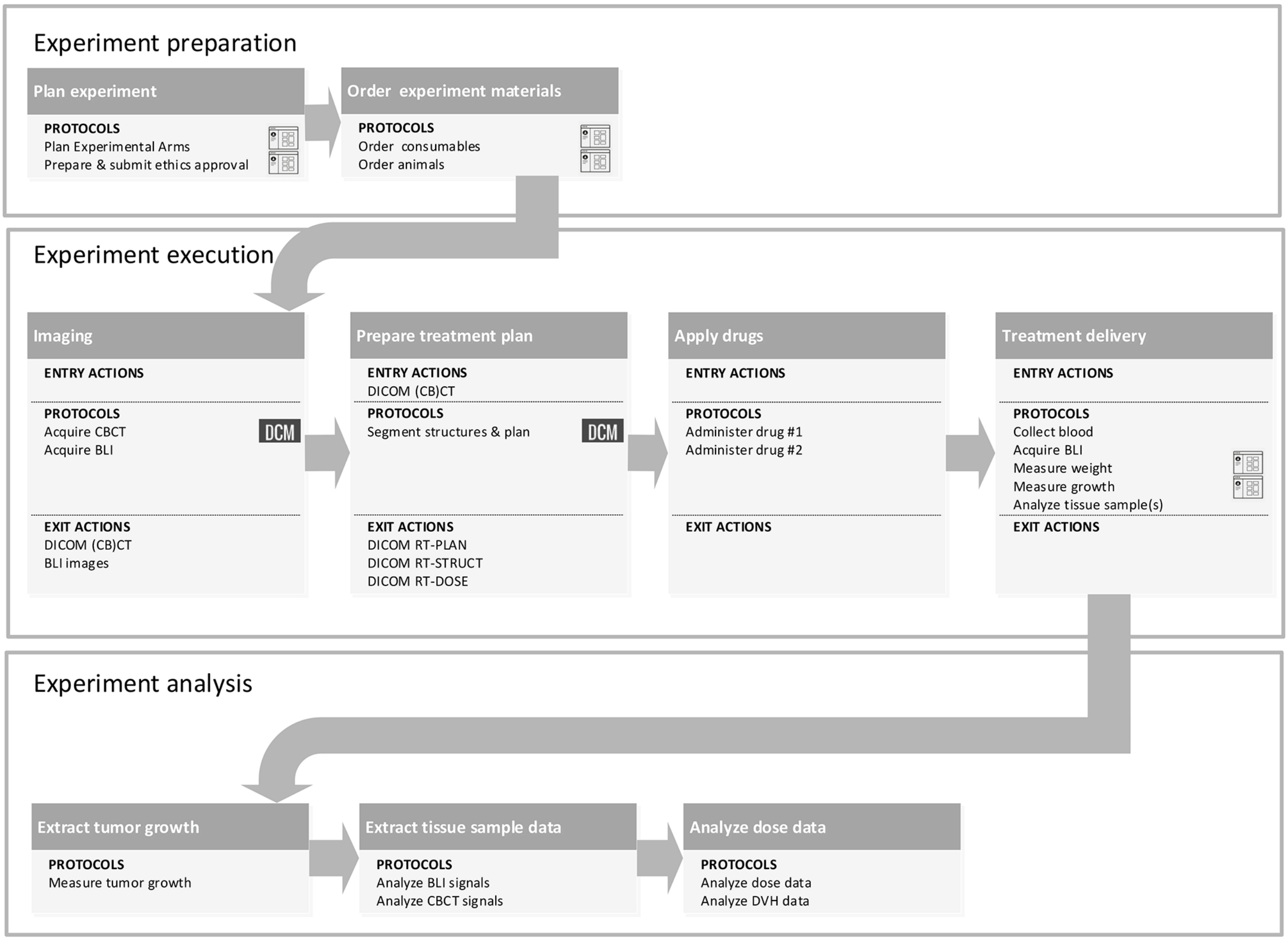
Example of use of the data management system

Figure 5 shows an example of all the steps involved in a typical experiment involving stages of preparation, execution and analysis. Details and complexity can vary substantially for different experiments. This kind of design can be made with, *e.g.* the eclipse sirius tool,[15] using an intuitive graphical user environment, easy click and drag tools, and stored templates. All the tasks generate actions, which take the user to various steps in the data management system. Once this template is filled, an entire management structure is generated, where the general idea is to support the user in all steps, ensuring, *e.g.* that no experimental arms are forgotten, that experimental materials and animals are ordered, and that administering of drugs of radiation is performed timely. A clear timeline would be generated, with all the tasks to be performed, indicating the persons who will perform tasks, and this could be linked to electronic agendas of the users.

DISCUSSION

In establishing radiotherapy practice, radiobiology research has not yet played a major role. Radiobiology has mostly yielded fundamental insights in radiation action on living tissue but it has not yet led to major translational breakthroughs.[4,16,17] This may change now with the increasing availability of novel tumor animal models (often orthotopic) and image-guided precision irradiation platforms. These platforms have seen an exponential growth in use, from less than 10 users a decade ago when the field started, to more than 100 now and is still rapidly growing. The development of these platforms has focused mainly on the radiation targeting and imaging capabilities. Another essential part of the technology to enable the clinical translation process is a data management system. This can be designed in analogy with clinical radiotherapy data management systems but there are also essential differences such as the focus on the cohort

Figure 5. Example of the workflow of the preparation, execution and analysis of an image-guided radiation experiment integrated with the data management system. Details can vary per experiment but the three main blocks will usually be present. The term protocol refers to actions/activities within the subtasks in each block, which should be standardized. The example workflow starts from the experiment definition, runs through the experiment execution with the acquisition of CBCT images and BLI to the treatment delivery and finishes with the experiment analysis analyzing the CBCTs, BLIs and DVHs. BLI, bioluminescence images; CBCT, cone-beam CT; DVHs, dose volume histograms.



instead of the single patient, the wide range of experimental techniques (imaging, genomics, histology,…) and the vast amount and versatility of the data generated. Whereas for a typical radiotherapy patient about 10 gigabyte of data is handled, for a preclinical cohort this can run into the terabytes, especially when raw imaging data is stored for further offline analysis.

The potentially large amount of data generated in preclinical studies also requires powerful data analysis methods, such as data mining and radiomics image analysis. To integrate seamlessly the data analysis with the setup and execution of preclinical experiments, a platform such as the one proposed here is required. The design should heavily focus on efficiency and user-friendliness, also taking into account that many users will not be experts in informatics systems, unlike in a clinical radiotherapy environment where full support of these systems is available.

In radiotherapy, standardization has benefitted clinical practice and internal and external communication. It took many years to reach the current standards, and the design of a preclinical platform can greatly profit from the efforts in radiotherapy informatics. The benefits of a preclinical data management system are:

- A centralized point of communication with complete transparency
- Coherent study object data operations
- Centralized storage of study & cohort records
- Clear guidance through preparation, execution and analysis of experiments
- Leveraging of information from different data sources
- Increased performance and operation efficiency
- Data standardization
- Easy communication with other institutes

One of the bigger challenges remains the translation of preclinical findings into clinical trials and knowledge.[17] However, some roadblocks and issues have been reported for successful translation of pre-clinical research.[16,18] These roadblocks and issues encompass a fragmented infrastructure, incompatible databases and consistent data acquisition, to name just a few items. These are the more technical blocks and issues which might be solved by the proposed solution. The rules to enable translational research with these type of platforms still largely have to be discovered. However, the novel platforms, including a powerful data management system, may greatly facilitate this process.

CONCLUSION

It can be concluded that a data management system such as the one proposed here, and largely inspired by current radiotherapy data management systems, will benefit preclinical radiation research. There are many similarities between the clinical and preclinical situation, but there are also essential differences. An important aim of a preclinical data management system should be to improve experimental workflow but the final aim should be to support translation of preclinical findings into clinical trials.

REFERENCES

1. Choplin RH, Boehme JM, Maynard CD. Picture archiving and communication systems: an overview. *Radiographics* 1992; **12**: 127–9. doi: https://doi.org/10.1148/radiographics.12.1.1734458
2. DICOM (Digital Imaging and Communications in Medicine) standard. 2018; National Electrical Manufacturers Association. Available from: www.dicom.nema.org.
3. Verhaegen F, Granton P, Tryggestad E. Small animal radiotherapy research platforms. *Phys Med Biol* 2011; **56**: R55–R83. doi: https://doi.org/10.1088/0031-9155/56/12/R01
4. Tillner F, Thute P, Bütof R, Krause M, Enghardt W. Pre-clinical research in small animals using radiotherapy technology--a bidirectional translational approach. *Z Med Phys* 2014; **24**: 335–51. doi: https://doi.org/10.1016/j.zemedi.2014.07.004
5. Bert C, Engenhart-Cabillic R, Durante M. Particle therapy for noncancer diseases. *Med Phys* 2012; **39**: 1716–27. doi: https://doi.org/10.1118/1.3691903
6. Granton PV, Dubois L, van Elmpt W, van Hoof SJ, Lieuwes NG, De Ruysscher D, et al. A longitudinal evaluation of partial lung irradiation in mice by using a dedicated image-guided small animal irradiator. *Int J Radiat Oncol Biol Phys* 2014; **90**: 696–704. doi: https://doi.org/10.1016/j.ijrobp.2014.07.004
7. De Ruysscher D, Granton PV, Lieuwes NG, van Hoof S, Wollin L, Weynand B, et al. Nintedanib reduces radiation-induced microscopic lung fibrosis but this cannot be monitored by CT imaging: A preclinical study with a high precision image-guided irradiator. *Radiother Oncol* 2017; **124**: 482–7. doi: https://doi.org/10.1016/j.radonc.2017.07.014
8. Belcaid Z, Phallen JA, Zeng J, See AP, Mathios D, Gottschalk C, et al. Focal radiation therapy combined with 4-1BB activation and CTLA-4 blockade yields long-term survival and a protective antigen-specific memory response in a murine glioma model. *PLoS One* 2014; **9**: e101764. doi: https://doi.org/10.1371/journal.pone.0101764
9. Yahyanejad S, Granton PV, Lieuwes NG, Gilmour L, Dubois L, Theys J, et al. Complementary use of bioluminescence imaging and contrast-enhanced micro-computed tomography in an orthotopic brain tumor model. *Mol Imaging* 2014; **13**: 7290.2014.00038. doi: https://doi.org/10.2310/7290.2014.00038
10. Gutierrez S, Descamps B, Vanhove C. MRI-Only based radiotherapy treatment planning for the rat brain on a small animal radiation research platform (SARRP). *PLoS One* 2015; **10**: e0143821. doi: https://doi.org/10.1371/journal.pone.0143821
11. KlingerC. CTIIP Primer. 2016; National Cancer Institute Wiki. Available from:https://wiki.nci.nih.gov/display/CTIIP/CTIIP+Primer#CTIIPPrimer-Supplement187DataElements.
12. DICOM Supplement 187 - Preclinical Small Animal Imaging Acquisition Context. 2015; National Electrical Manufacturers Association - DICOM standard. Available from:ftp://medical.nema.org/medical/dicom/final/sup187_ft_preclinicalanimalacquisitioncontext.pdf.
13. Aerts J. Towards a single data exchange standard for use in healthcare and in clinical research. *Stud Health Technol Inform* 2018; **248**: 55–63.
14. van Hoof SJ, Granton PV, Verhaegen F. Development and validation of a treatment planning system for small animal radiotherapy: SmART-Plan. *Radiother Oncol* 2013; **109**: 361–6. doi: https://doi.org/10.1016/j.radonc.2013.10.003
15. Modelio 3.7.1. 2018; https://www.modelio.org/
16. Liu FF, workshop participants, Okunieff P, Bernhard EJ, Stone HB, Yoo S, et al. Lessons learned from radiation oncology clinical trials. *Clin Cancer Res* 2013; **19**: 6089–100. doi: https://doi.org/10.1158/1078-0432.CCR-13-1116
17. Koontz BF, Verhaegen F, De Ruysscher D. Tumour and normal tissue radiobiology in mouse models: how close are mice to mini-humans? *Br J Radiol* 2017; **90**: 20160441. doi: https://doi.org/10.1259/bjr.20160441
18. Trochim W, Kane C, Graham MJ, Pincus HA. Evaluating translational research: a process marker model. *Clin Transl Sci* 2011; **4**: 153–62. doi: https://doi.org/10.1111/j.1752-8062.2011.00291.x